

Designing dose finding studies with an active control for exponential families

Holger Dette, Katrin Kettelhake

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

e-mail: holger.dette@ruhr-uni-bochum.de

Frank Bretz

Statistical Methodology

Novartis Pharma AG

4002 Basel, Switzerland

e-mail: frank.bretz@novartis.com

November 3, 2014

Abstract

In a recent paper Dette et al. (2014) introduced optimal design problems for dose finding studies with an active control. These authors concentrated on regression models with normal distributed errors (with known variance) and the problem of determining optimal designs for estimating the smallest dose, which achieves the same treatment effect as the active control. This paper discusses the problem of designing active-controlled dose finding studies from a broader perspective. In particular, we consider a general class of optimality criteria and models arising from an exponential family, which are frequently used analyzing count data. We investigate under which circumstances optimal designs for dose finding studies including a placebo can be used to obtain optimal designs for studies with an active control. Optimal designs are constructed for several situations and the differences arising from different distributional assumptions are investigated in detail. In particular, our results are applicable for constructing optimal experimental designs to analyze active-controlled dose finding studies with discrete data, and we illustrate the efficiency of the new optimal designs with two recent examples from our consulting projects.

Keywords and Phrases: optimal designs, dose response, dose estimation, active control

1 Introduction

Dose finding studies are an important tool to investigate the effect of a compound on a response of interest and have numerous applications in various fields such as medicine, biology or toxicology. They are of particular importance in pharmaceutical drug development

because marketed doses have to be safe and provide clinically relevant efficacy [see Ruberg (1995); Ting (2006)]. Most of the literature on statistical methodology for analyzing dose response studies include placebo as a control group [see Pinheiro et al. (2006); Bretz et al. (2008), among others]. Numerous authors have worked on the problem of determining optimal designs for dose response experiments with a placebo group because the application of efficient designs can substantially increase the accuracy of statistical analysis [see Zhu and Wong (2000); Fedorov and Leonov (2001); Krewski et al. (2002); Wu et al. (2005); Dragalin et al. (2007); Miller et al. (2007); Bornkamp et al. (2011), among many others].

However, dose response studies including a marketed drug as an active control are becoming more popular, especially in preparation for an active-controlled confirmatory non-inferiority trial where the use of placebo may be unethical. Thus, considerable interest on active-controlled studies has emerged, as documented through the release of several related guidelines by regulatory agencies [see ICH (1994), EMEA (2006, 2011), EMEA (2005)]. Recently Helms et al. (2014a) investigated the finite sample properties of maximum likelihood estimates of the target dose in an active-controlled study, which achieves the same treatment effect as the active control and Helms et al. (2014b) studied nonparametric estimates for this quantity.

Despite of these important applications, to our best knowledge, optimal design problems for active-controlled dose finding studies have only been considered in one paper so far [Dette et al. (2014)]. These authors investigated optimal designs for estimating the target dose under the assumption of a normal distribution with known variances. In particular, they demonstrated the superiority of the optimal designs compared to standard designs used in pharmaceutical practice. However, this work is restricted to normal distributed responses with known variances and a special c -optimality criterion and obviously the designs derived in their paper are not necessarily useful for other applications. Therefore the goal of the present paper is to investigate optimal design problems for dose finding studies with an active control from a more general perspective. A first objective is to consider a general class of optimality criteria. Second, as it will be pointed out in the following paragraph, in many dose finding trials with an active control the assumption of normal distributed responses is hard to justify, and we consider exponential families for modeling the distribution of the responses of the new drug and the active control. This allows in particular to design experiments for controlled studies with discrete data as they have appeared in the consulting projects described in the next paragraph. Third, even if the assumption of a normal distribution is justifiable, we will demonstrate that the estimation of the variances has a nontrivial effect on the optimal designs for an active-controlled study.

The research in the present paper is motivated by two clinical trial examples where the assumption of normal distributed responses made by Dette et al. (2014) is hard to justify. The first example refers to a 24-week, dose-ranging, Phase II study in patients with gouty arthritis to determine the target dose of a compound in preventing signs and symptoms of

flares in chronic gout patients starting allopurinol therapy. The study population consists of male and female patients (age 18 – 80 years) diagnosed with chronic gout as defined by the American College of Rheumatology preliminary criteria (ACR) and willing to either initiate allopurinol therapy or having just initiated allopurinol therapy within less than one month. Approximately 500 patients are screened in order to randomize 440 patients in approximately 100 centers worldwide. Patients who meet the entry criteria are randomized to receive either the active comparator or a specific dose of the new compound. The primary endpoint is the number of flares occurring per subject within 16 weeks of randomization, which are modeled using a negative binomial distribution for all treatment arms, where the corresponding probability is modeled by a dose-response relationship between the (single) dose groups of the new compound and by a constant parameter for the comparator.

The second example is a Phase IIb, multicenter, randomized, double-blind, active-controlled dose-finding study in the treatment of acute migraine, as measured by the percentage of patients reporting pain freedom at two hours post-dose. Approximately 500 patients are randomized worldwide. Patients who meet the entry criteria are randomized to receive either the active comparator or one dose of the new compound. Once the dose of the new compound is selected, Phase III studies are conducted to evaluate further the efficacy and safety of the new compound in the targeted patient population.

In Section 2 we give an introduction to optimal design theory for models with an active control under general distributional assumptions. In particular we present results, which relate optimal designs for dose finding studies with a placebo group to optimal designs for models with an active control. This methodology is used in Section 3 to construct D -optimal designs for dose finding studies with an active control. In Section 4 we consider the optimal design problem for estimating the smallest dose, which achieves the same treatment effect as the active control. In both sections we investigate the effect of the distributional assumption on the resulting optimal designs. In particular, we show that different distributional assumptions (as for example a normal or Poisson distribution used to model continuous or discrete data) leads to substantial changes in the structure of the optimal designs. The Appendix contains the proofs of our main results.

For the sake of brevity this paper is restricted to locally optimal designs which require a-priori information about the unknown model parameters [see Chernoff (1953), Ford et al. (1992), Fang and Hedayat (2008)]. These designs can be used as benchmarks for commonly used designs. Moreover, locally optimal designs serve as basis for constructing optimal designs with respect to more sophisticated optimality criteria, which are robust against a misspecification of the unknown parameters [see Pronzato and Walter (1985) or Chaloner and Verdinelli (1995), Dette (1997), Imhof (2001) among others]. Following this line of research the methodology introduced in the present paper can be further developed to adress uncertainty in the preliminary information for the unknown parameters.

2 Modeling active-controlled dose finding studies using exponential families

Consider a clinical trial, where patients are treated either with an active control (a standard treatment administered at a fixed dose level) or with a new drug using different dose levels in order to investigate the corresponding dose response relationship. Given a total sample size N , we thus allocate n_1 and $n_2 = N - n_1$ patients to the two treatments. In addition, we determine the optimal number of different dose levels for the new drug, the dose levels themselves and the optimal number of patients allocated to each dose level to obtain the design of the experiment.

More formally, we assume that k different dose levels, say d_1, \dots, d_k , are chosen in a dose range, say $\mathcal{D} \subset \mathbb{R}_0^+$, for the new drug (the optimal number k and the dose levels will be determined by the choice of the design) and that at each dose level d_i the experimenter can investigate n_{1i} patients ($i = 1, \dots, k$), where $n_1 = \sum_{i=1}^k n_{1i}$ denotes the number of patients treated with the new drug. The optimal numbers n_{1i} , more precisely the optimal proportions n_{1i}/n_1 , will be determined by the choice of the design. The corresponding responses at dose level d_i are modeled as realizations of independent real valued random variables Y_{ij} ($j = 1, \dots, n_{1i}$, $i = 1, \dots, k$). Similarly, the responses of patients treated with the active control are modeled as realizations of independent real valued random variables Z_1, \dots, Z_{n_2} , where the two samples corresponding to the new drug and active control are assumed to be independent. For the statistical analysis we further assume that the random variables Z_j and Y_{ij} have distributions from an exponential family, where the distributions of the latter depend on the corresponding dose levels d_i , that is

$$f_1(y|d_i, \theta_1) := \frac{\partial P_{\theta_1}^{Y_{i1}}}{\partial \nu}(y) = \exp\{c_1^T(d_i, \theta_1)T_1(y) - b_1(d_i, \theta_1)\}h_1(y), \quad (2.1)$$

$$f_2(z|\theta_2) := \frac{\partial P_{\theta_2}^{Z_1}}{\partial \nu}(z) = \exp\{c_2^T(\theta_2)T_2(z) - b_2(\theta_2)\}h_2(z). \quad (2.2)$$

Here ν denotes a σ -finite measure on the real line, $\theta_1 \in \Theta_1 \subset \mathbb{R}^{s_1}$, $\theta_2 \in \Theta_2 \subset \mathbb{R}^{s_2}$ are unknown parameters and we use common terminology for exponential families [see for example Brown (1986)]. In particular, the functions $c_1 : \mathcal{D} \times \Theta_1 \rightarrow \mathbb{R}^{\ell_1}$, $b_1 : \mathcal{D} \times \Theta_1 \rightarrow \mathbb{R}$, $c_2 : \Theta_2 \rightarrow \mathbb{R}^{\ell_2}$ and $b_2 : \Theta_2 \rightarrow \mathbb{R}$ are assumed to be twice continuously differentiable where $\frac{\partial c_1}{\partial \theta_1}, \frac{\partial c_2}{\partial \theta_2} \neq 0$ and T_1 and T_2 denote ℓ_1 - and ℓ_2 - dimensional statistics defined on the corresponding sample spaces. Additionally, the functions h_1 and h_2 are assumed to be positive (and measurable).

Throughout this paper let κ be a variable indicating whether a patient receives the new drug ($\kappa = 0$) or the active control ($\kappa = 1$) and denote

$$\mathcal{X} = (\mathcal{D} \times \{0\}) \cup \{(C, 1)\} \quad (2.3)$$

as the design space of the experiment, where \mathcal{D} is the dose range for the new drug, C the dose level of the active control and the second component of an experimental condition $(d, \kappa) \in \mathcal{X}$ determines the treatment ($\kappa = 0, 1$). Straightforward calculation shows that the Fisher information at the point $(d, \kappa) \in \mathcal{X}$ is given by the matrix

$$I((d, \kappa), \theta) = \begin{pmatrix} I\{\kappa = 0\}I_1(d, \theta_1) & \mathbf{0} \\ \mathbf{0} & I\{\kappa = 1\}I_2(\theta_2) \end{pmatrix}, \quad (2.4)$$

where $\mathbf{0}$ denotes a matrix of appropriate dimension with all entries equal to 0, $\theta = (\theta_1^T, \theta_2^T)^T \in \Theta_1 \times \Theta_2 \subset \mathbb{R}^{s_1+s_2}$ is the vector of all parameters, $I\{\kappa = 0\}$ is the indicator function of the event $\{\kappa = 0\}$ and the matrices I_1 and I_2 are the Fisher information matrices of the two models (2.1) and (2.2), that is

$$\begin{aligned} I_1(d, \theta_1) &= \mathbb{E} \left[\left(\frac{\partial}{\partial \theta_1} \log f_1(Y|d, \theta_1) \right) \left(\frac{\partial}{\partial \theta_1} \log f_1(Y_{i1}|d_i, \theta_1) \right)^T \right] \\ I_2(\theta_2) &= \mathbb{E} \left[\left(\frac{\partial}{\partial \theta_2} \log f_2(Z|\theta_2) \right) \left(\frac{\partial}{\partial \theta_2} \log f_2(Z_1|\theta_2) \right)^T \right]. \end{aligned} \quad (2.5)$$

where the random variables Y and Z have densities $f_1(y|d, \theta_1)$ and $f_2(z|\theta_2)$ defined by (2.1) and (2.2), respectively. Note that the Fisher information in (2.4) is block diagonal because of the independence of the samples, as different patients are either treated with the new drug or the active control. The following examples illustrate the general terminology.

Example 2.1 In order to demonstrate the different structures of the Fisher information arising from different distributions of the exponential family we consider several examples.

- (a) Dette et al. (2014) investigated normal distributed responses with known variances σ_1^2 and σ_2^2 for the new drug and the active control, respectively. For the expectation of the response of the new drug at dose level d they assumed a nonlinear regression model, say $\eta(d, \vartheta)$, where $\vartheta = (\vartheta_0, \dots, \vartheta_s)$, while it is assumed to be equal to μ for the active control. If the variances are not known and have to be estimated from the data, we have $\theta_1 = (\vartheta_0, \dots, \vartheta_s, \sigma_1^2)$, $\theta_2 = (\mu, \sigma_2^2)$ for the parameters in models (2.1) and (2.2), respectively. Standard calculations show that the Fisher information at a point $(d, \kappa) \in \mathcal{X}$ is given by (2.4), where

$$I_1(d, \theta_1) = \begin{pmatrix} \frac{1}{\sigma_1^2} \left(\frac{\partial}{\partial \vartheta} \eta(d, \vartheta) \right) \left(\frac{\partial}{\partial \vartheta} \eta(d, \vartheta) \right)^T & \mathbf{0} \\ \mathbf{0} & \frac{1}{2\sigma_1^4} \end{pmatrix}, \quad I_2(\theta_2) = \begin{pmatrix} \frac{1}{\sigma_2^2} & 0 \\ 0 & \frac{1}{2\sigma_2^4} \end{pmatrix}. \quad (2.6)$$

- (b) As motivated by the examples in Section 1, it might be more reasonable to consider a different distribution than a normal distribution in (2.1) and (2.2) to model discrete data. Assume, for example, a negative binomial distribution with parameter $r_1 \in \mathbb{N}$ for

the number of failures and a function $\pi(d, \theta_1) \in (0, 1)$ for the probability of a success of the new drug (at dose level d) and parameters $r_2 \in \mathbb{N}, \mu \in (0, 1)$ for the active control. Then we have $\theta_2 = \mu$ and the Fisher information matrix is given by (2.4), where

$$I_1(d, \theta_1) = \frac{r_1(\frac{\partial}{\partial \theta_1} \pi(d, \theta_1))(\frac{\partial}{\partial \theta_1} \pi(d, \theta_1))^T}{\pi^2(d, \theta_1)(1 - \pi(d, \theta_1))}, \quad I_2(\theta_2) = \frac{r_2}{\mu^2(1 - \mu)}. \quad (2.7)$$

Here the parameters $r_1, r_2 \in \mathbb{N}$ for the number of failures are assumed to be known.

- (c) Alternatively, in the case of a binary response we may use a Bernoulli distribution, where $\pi(d, \theta_1) \in (0, 1)$ and $\mu \in (0, 1)$ denote the probability of success for the new drug and the active control, respectively. In this case we have $\theta_2 = \mu$ and the Fisher information matrix is given by (2.4), where

$$I_1(d, \theta_1) = \frac{(\frac{\partial}{\partial \theta_1} \pi(d, \theta_1))(\frac{\partial}{\partial \theta_1} \pi(d, \theta_1))^T}{\pi(d, \theta_1)(1 - \pi(d, \theta_1))}, \quad I_2(\theta_2) = \frac{1}{\mu(1 - \mu)}.$$

- (d) For a Poisson distribution, the parameters for the distribution of the responses corresponding to the new drug and the active control are given by a function of the dose level, say $\lambda(d, \theta_1) > 0$, and a parameter $\mu > 0$, respectively. In this case we have $\theta_2 = \mu$ and the Fisher information matrix is given by (2.4), where the two non-vanishing blocks are defined by

$$I_1(d, \theta_1) = \frac{(\frac{\partial}{\partial \theta_1} \lambda(d, \theta_1))(\frac{\partial}{\partial \theta_1} \lambda(d, \theta_1))^T}{\lambda(d, \theta_1)}, \quad I_2(\theta_2) = \frac{1}{\mu}. \quad (2.8)$$

Throughout this paper we consider approximate designs in the sense of Kiefer (1974), which are defined as probability measures with finite support on the design space \mathcal{X} in (2.3). Therefore, an experimental design is given by

$$\xi = \begin{pmatrix} (d_1, 0) & \dots & (d_k, 0) & (C, 1) \\ w_1 & \dots & w_k & w_{k+1} \end{pmatrix}, \quad (2.9)$$

where w_1, \dots, w_{k+1} are positive weights, such that $\sum_{i=1}^{k+1} w_i = 1$. Here, w_i denotes the relative proportion of patients treated at dose level d_i ($i = 1, \dots, k$) or the active control ($i = k + 1$). If N observations can be taken, a rounding procedure is applied to obtain integers n_{1i} ($i = 1, \dots, k$) and n_2 from the not necessarily integer valued quantities $w_i N$ ($i = 1, \dots, k + 1$) [see Pukelsheim and Rieder (1992)]. Thus, the experimenter assigns n_{11}, \dots, n_{1k} and n_2 patients to the dose levels d_1, \dots, d_k of the new drug and the active control, respectively. In the following discussion we will determine optimal designs, which also optimize the number k of different dose levels. It turns out that for the models considered

here the optimal designs usually allocate observations at less than 5 dose levels. Note that in practice the number k of different dose levels is in the range of 4 – 7 and rarely larger than 10.

The information matrix of an approximate design ξ of the form (2.9) is defined by the $(s_1 + s_2) \times (s_1 + s_2)$ matrix

$$M(\xi, \theta) = \int_{\mathcal{X}} I((d, \kappa), \theta) d\xi(d, \kappa) = \begin{pmatrix} (1 - \omega_{k+1})M_1(\tilde{\xi}, \theta_1) & \mathbf{0} \\ \mathbf{0} & \omega_{k+1}I_2(\theta_2) \end{pmatrix} \quad (2.10)$$

Here, the $s_1 \times s_1$ matrix $M_1(\tilde{\xi}, \theta_1)$ and the $s_2 \times s_2$ matrix $I_2(\theta_2)$ are given by

$$M_1(\tilde{\xi}, \theta_1) = \int_{\mathcal{D}} I_1(d, \theta_1) d\tilde{\xi}(d), \quad (2.11)$$

and (2.5), respectively, and

$$\tilde{\xi} = \begin{pmatrix} d_1 & \dots & d_k \\ \tilde{w}_1 & \dots & \tilde{w}_k \end{pmatrix} \quad (2.12)$$

denotes the design (on the design space \mathcal{D}) for the new drug, which is induced by the design ξ in (2.9) defining the weights $\tilde{w}_i = \frac{w_i}{1 - w_{k+1}}$, $i = 1, \dots, k$.

If observations are taken according to an approximate design it can be shown (assuming standard regularity conditions) that the maximum likelihood estimators $\hat{\theta}_1, \hat{\theta}_2$ in models (2.1) and (2.2) are asymptotically normal distributed, that is

$$\sqrt{N}((\hat{\theta}_1^T, \hat{\theta}_2^T)^T - (\theta_1^T, \theta_2^T)^T) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, M^{-1}(\xi, \theta))$$

as $N \rightarrow \infty$, where the symbol $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution. Dette et al. (2008) considered dose finding studies including a placebo group and showed by means of a simulation study that the approximation of the variance of $\hat{\theta} = (\hat{\theta}_1^T, \hat{\theta}_2^T)^T$ by $\frac{1}{N}M^{-1}(\xi, \theta)$ is satisfactory for total sample sizes larger than 25. As typical clinical dose finding trials have sample sizes in the range of 200 – 300 [see for example Bornkamp et al. (2007)], it is reasonable to use this approximation also for active-controlled studies. Consequently, optimal designs maximize an appropriate functional of the information matrix defined in (2.10).

In order to discriminate between competing designs we consider in this paper Kiefer's ϕ_p -criteria [see Kiefer (1974) or Pukelsheim (2006)]. To be precise, let $p \in [-\infty, 1)$ and $K \in \mathbb{R}^{(s_1+s_2) \times t}$ denote a matrix of full column rank t . Then a design ξ^* is called locally ϕ_p -optimal for estimating the linear combination $K^T\theta$ in a dose response model with an active control, if $K^T\theta$ is estimable by the design ξ^* , that is, $K^T\theta \in \text{Range}(M(\xi^*, \theta))$, and ξ^* maximizes the functional

$$\phi_p(\xi) = \left(\frac{1}{t} \text{tr}(K^T M^{-1}(\xi, \theta) K)^{-p} \right)^{\frac{1}{p}} \quad (2.13)$$

among all designs for which $K^T\theta$ is estimable, where $\text{tr}(A)$ and A^- denote the trace and a generalized inverse of the matrix A , respectively. Note that the cases $p = 0$ and $p = -\infty$ correspond to the D - and E -optimality criterion, that is $\phi_0(\xi) = \det(K^T M^-(\xi, \theta) K)^{-\frac{1}{t}}$ and $\phi_{-\infty}(\xi) = \lambda_{\min}((K^T M^-(\xi, \theta) K)^{-1})$. An application of the general equivalence theorem [see Pukelsheim (2006), chapter 7.19 and 7.21, respectively] to the situation considered in this paper yields immediately the following result.

Lemma 2.1 *If $p \in (-\infty, 1)$, a design ξ^* with $K^T\theta \in \text{Range}(M(\xi^*, \theta))$ is locally ϕ_p -optimal for estimating the linear combination $K^T\theta$ in a dose response model with an active control if and only if there exists a generalized inverse G of the information matrix $M(\xi^*, \theta)$, such that the inequality*

$$\text{tr}(I((d, \kappa), \theta) G K (K^T M^-(\xi^*, \theta) K)^{-p-1} K^T G^T) - \text{tr}(K^T M^-(\xi^*, \theta) K)^{-p} \leq 0 \quad (2.14)$$

holds for all $(d, \kappa) \in \mathcal{X}$. If $p = -\infty$, a design ξ^ with $K^T\theta \in \text{Range}(M(\xi^*, \theta))$ is locally $\phi_{-\infty}$ -optimal for estimating the linear combination $K^T\theta$ if and only if there exist a generalized inverse G of the information matrix $M(\xi^*, \theta)$ and a nonnegative definite matrix $E \in \mathbb{R}^{t \times t}$ with $\text{tr}(E) = 1$, such that the inequality*

$$\text{tr}(I((d, \kappa), \theta) G K (K^T M^-(\xi^*, \theta) K)^{-1} E (K^T M^-(\xi^*, \theta) K)^{-1} K^T G^T) - \lambda_{\min}((K^T M^-(\xi^*, \theta) K)^{-1}) \leq 0 \quad (2.15)$$

holds for all $(d, \kappa) \in \mathcal{X}$. Moreover, there is equality in (2.14) ($p > -\infty$) and (2.15) ($p = -\infty$) for all support points of the design ξ^ .*

In the following discussion we assume that either $p = -1$ or that the matrix K is a block matrix of the form

$$K = \begin{pmatrix} K_{11} & 0 \\ 0 & K_{22} \end{pmatrix} \in \mathbb{R}^{(s_1+s_2) \times (t_1+t_2)} \quad (2.16)$$

with elements $K_{11} \in \mathbb{R}^{s_1 \times t_1}$, $K_{22} \in \mathbb{R}^{s_2 \times t_2}$, $t_1 + t_2 = t$. Roughly speaking, the choice $p = -1$ or a blockdiagonal structure of the matrix K in (2.16) leads to a separation of the parameters from models (2.1) and (2.2) in the corresponding optimality criterion. As a consequence optimal designs for dose finding studies with an active control can be obtained from optimal designs for dose finding studies including a placebo group, which maximize the criterion

$$\tilde{\phi}_p(\tilde{\xi}) = \left(\frac{1}{t_1} \text{tr}(K_{11}^T M_1^-(\tilde{\xi}, \theta_1) K_{11})^{-p} \right)^{\frac{1}{p}} \quad (2.17)$$

in the class of all designs $\tilde{\xi}$ for which $K_{11}^T \theta_1$ is estimable, i.e. $K_{11}^T \theta_1 \in \text{Range}(M_1(\tilde{\xi}, \theta_1))$. Throughout this paper these designs are called $\tilde{\phi}_p$ -optimal for estimating the parameter $K_{11}^T \theta_1$ in the dose response model (2.1). The proof can be found in the Appendix.

Theorem 2.1 Assume $p \in [-\infty, 1)$, that the matrix K is given by (2.16) and that

$$\tilde{\xi}_p^* = \begin{pmatrix} d_1^* & \dots & d_k^* \\ \tilde{w}_1^* & \dots & \tilde{w}_k^* \end{pmatrix} \quad (2.18)$$

is a locally $\tilde{\phi}_p$ -optimal design for estimating $K_{11}^T \theta_1$ in the dose response model (2.1). Then the design

$$\xi_p^* = \begin{pmatrix} (d_1^*, 0) & \dots & (d_k^*, 0) & (C, 1) \\ w_1^* & \dots & w_k^* & w_{k+1}^* \end{pmatrix}$$

is locally ϕ_p -optimal for estimating $K^T \theta$ in the dose response model with an active control, where the weights are given by

$$w_{k+1}^* = \frac{1}{1 + \rho_p}, \quad w_i^* = \frac{\rho_p}{1 + \rho_p} \tilde{w}_i^* \quad (i = 1, \dots, k), \quad (2.19)$$

and

$$\rho_p = \frac{(\text{tr}((K_{22}^T I_2^-(\theta_2) K_{22})^{-p}))^{1/(p-1)}}{(\text{tr}((K_{11}^T M_1^-(\tilde{\xi}_p^*, \theta_1) K_{11})^{-p}))^{1/(p-1)}} \quad (2.20)$$

(the case $p = -\infty$ is interpreted as the corresponding limit).

In the case $p = -1$ a more general statement is available without the restriction to block matrices of the form (2.16). The proof is obtained by similar arguments as presented in the proof of Theorem 2.1 and therefore omitted.

Theorem 2.2 Assume that $K^T = (K_{11}^T, K_{22}^T) \in \mathbb{R}^{t \times (s_1 + s_2)}$ with $K_{11}^T \in \mathbb{R}^{t \times s_1}$, $K_{22}^T \in \mathbb{R}^{t \times s_2}$ and let $\tilde{\xi}_{-1}^*$ denote the $\tilde{\phi}_{-1}$ -optimal design for estimating the parameter $K_{11}^T \theta_1$ in the dose response model (2.1). Then the design ξ_{-1}^* defined in Theorem 2.1 is locally ϕ_{-1} -optimal for estimating $K^T \theta$ in the dose response model with an active control.

The final result of this section considers the special case $p = 0$. The result is a direct consequence of Theorem 2.1 considering the limit $p \rightarrow 0$ and observing that the quantity ρ_p defined in (2.20) satisfies $\lim_{p \rightarrow 0} \rho_p = \frac{t_1}{t_2}$.

Corollary 2.1 Assume that the matrix K is given by (2.16) and let $\tilde{\xi}_0^*$ denote the locally D -optimal design of the form (2.18) for estimating the parameter $K_{11}^T \theta_1$ in the dose response model (2.1), which maximizes $\det((K_{11}^T M_1^-(\tilde{\xi}, \theta_1) K_{11})^{-1})$ in the class of all designs for which $K_{11}^T \theta_1$ is estimable. Then the design

$$\xi_\theta^* = \begin{pmatrix} (d_1^*, 0) & \dots & (d_k^*, 0) & (C, 1) \\ \frac{t_1}{t_1 + t_2} \tilde{w}_1^* & \dots & \frac{t_1}{t_1 + t_2} \tilde{w}_k^* & \frac{t_2}{t_1 + t_2} \end{pmatrix} \quad (2.21)$$

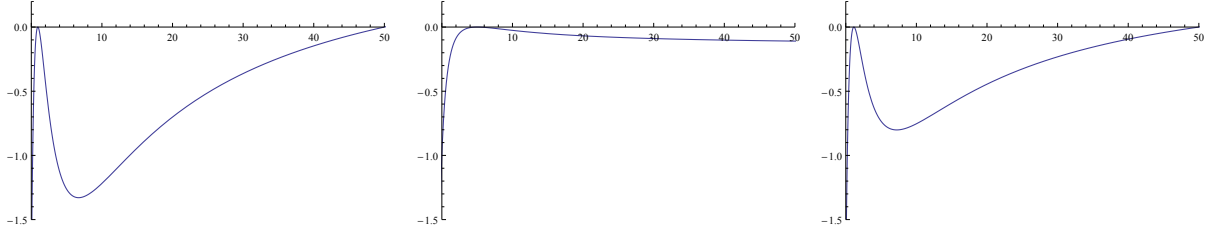


Figure 1: *The inequality (2.14) of the equivalence theorem. Left panel: D -optimal design for estimating $K^T\theta$ in the dose response model with an active control. Middle panel and right panel: optimal design for prediction and D -optimal design in the model (2.1).*

is locally D -optimal for estimating the parameter $K^T\theta$ in the dose response model with an active control.

Remark 2.1 The assumption of a block matrix K in Theorem 2.1 and Corollary 2.1 can not be omitted. Consider for example the case of a binomial distribution and a Michaelis-Menten model $\pi(d, \theta_1) = \frac{\vartheta_1 d}{\vartheta_2 + d}$ for the dose response relationship of the new drug. Assume that one is interested in two functionals of the model parameters: (i) The estimation of the distance between the effect of the active control, say θ_2 , and the effect of the new drug at a special dose level d_0 , i.e. $\pi(d_0, \theta_1)$, and (ii) the difference between θ_2 and the maximum effect of the new drug, i.e. ϑ_1 . In this case the matrix K is given by

$$K = \begin{pmatrix} -\frac{d_0}{\vartheta_2 + d_0} & \frac{\vartheta_1 d_0}{(\vartheta_2 + d_0)^2} & 1 \\ -1 & 0 & 1 \end{pmatrix}^T.$$

Consider exemplarily the choice $\mathcal{D} = [0, 50]$, $\vartheta_1 = 0.5$, $\vartheta_2 = 2$, $\theta_2 = 0.4$ and $d_0 = 5$. The locally D -optimal design for estimating $K^T\theta$ in the dose response model with an active control allocates 36% and 32% of the patients to the dose levels 0.93 and 50 of the new drug and 32% of the patients to the active control, respectively. The corresponding function (2.14) of the equivalence theorem is shown in the left panel of Figure 1 for the case $\kappa = 0$. In the case $\kappa = 1$ this function reduces to the constant 0 for all $d \in \mathcal{D}$. In the situation where no active control is available one could look at the problem designing the experiment for a most efficient estimation of $\pi(d_0, \theta_1)$. This corresponds to the matrix $K_{11} = (\frac{d_0}{\vartheta_2 + d_0}, -\frac{\vartheta_1 d_0}{(\vartheta_2 + d_0)^2})^T$ and the locally optimal design is a one point design which treats 100% of the patients with the dose level $d_0 = 5$. On the other hand the locally D -optimal design for estimating θ_1 allocates 50% of the patients to each of the dose levels 1.15 and 50. The corresponding inequalities of the equivalence theorem are shown in the middle and right panel of Figure 1. Obviously the locally D -optimal design for estimating $K^T\theta$ in the dose response model with an active control can not be derived from these designs and an assumption of the type (2.16) is in fact necessary to obtain Theorem 2.1.

3 D -optimal designs for the Michaelis-Menten and EMAX model

In this section we determine some D -optimal designs for dose finding studies with an active control under different distributional assumptions. We assume that the dependence on the dose of the new drug is either described by the Michaelis-Menten model $\frac{\vartheta_1 d}{\vartheta_2 + d}$, or the EMAX model $\vartheta_0 + \frac{\vartheta_1 d}{\vartheta_2 + d}$, where the dose range is given by the interval $[L, R] \subset \mathbb{R}_0^+$. These models are widely used when investigating the dose response relationship of a new compound, such as a medicinal drug, a fertilizer, or an environmental toxin. Note that in the case where the function describes a probability, one requires some restrictions on the parameters. For example, if $\pi(d, \theta_1) = \frac{\vartheta_1 d}{\vartheta_2 + d}$ is the probability of a success for the negative binomial distribution in Example 2.1(b), we implicitly assume $\frac{\vartheta_1 R}{\vartheta_2 + R} < 1$ in the following discussion. In other models similar assumptions have to be made and we do not mention these restrictions explicitly for the sake of brevity. In the following, $x \vee y$ denotes the maximum of $x, y \in \mathbb{R}$.

Theorem 3.1 (*Michaelis-Menten model*)

- (a) *If the distributions of the responses corresponding to the new drug and active control are normal with parameters $(\frac{\vartheta_1 d}{\vartheta_2 + d}, \sigma_1^2)$ and (μ, σ_2^2) , respectively, then the locally D -optimal design for the dose response model with an active control allocates 30% of the patients to each of the dose levels $L \vee \frac{\vartheta_2 R}{2\vartheta_2 + R}$ and R of the new drug and 40% to the active control.*
- (b) *In the case of negative binomial distributions with probabilities $\pi(d, \theta) = \frac{\vartheta_1 d}{\vartheta_2 + d}$ and μ the locally D -optimal design for the dose response model with an active control allocates 33. $\bar{3}$ % of the patients to each of the dose levels L and R of the new drug and 33. $\bar{3}$ % to the active control.*
- (c) *In the case of binomial distributions with probabilities $\pi(d, \theta) = \frac{\vartheta_1 d}{\vartheta_2 + d}$ and μ the locally D -optimal design for the dose response model with an active control allocates 33. $\bar{3}$ % of the patients to each of the dose levels $L \vee \frac{\vartheta_2 R + 3\vartheta_2^2 - \vartheta_2 \sqrt{9R^2 - 8R^2\vartheta_1 + 18R\vartheta_2 - 8R\vartheta_1\vartheta_2 + 9\vartheta_2^2}}{4\vartheta_1\vartheta_2 - 4R + 4R\vartheta_1 - 6\vartheta_2}$ and R of the new drug and 33. $\bar{3}$ % to the active control.*
- (d) *If Poisson distributions with parameters $\lambda(d, \theta_1) = \frac{\vartheta_1 d}{\vartheta_2 + d}$ and μ are used in (2.1) and (2.2), the locally D -optimal design for the dose response model with an active control allocates 33. $\bar{3}$ % of the patients to each of the dose levels $L \vee \frac{\vartheta_2 R}{3\vartheta_2 + 2R}$ and R of the new drug and 33. $\bar{3}$ % to the active control.*

The proof of Theorem 3.1 is a direct consequence of Corollary 2.1, if the locally D -optimal designs for model (2.1) are known. For example, in the case of a normal distribution it

follows from Rasch (1990) that the D -optimal design for the Michaelis-Menten model has equal masses at the points $L \vee \frac{\vartheta_2 R}{2\vartheta_2 + R}$ and R and Corollary 2.1 yields part (a) of Theorem 3.1. In the other cases the D -optimal designs for model (2.1) are not known and the proof can be found in the Appendix.

It is also worthwhile to note that the differences of the D -optimal designs derived under different distributional assumptions can be substantial. For example if the design space is $[0, R]$ with a large right boundary R , the non trivial dose level for the new drug is approximately ϑ_2 and $\vartheta_2/2$ under the assumption of a normal and Poisson distribution, respectively. We will now give the corresponding results for the EMAX model. The proof follows by similar arguments as given in the proof of Theorem 3.1 and is therefore omitted.

Theorem 3.2 (*EMAX-model*)

- (a) *If the distribution of responses corresponding to the new drug and active control are normal distributions with parameters $(\vartheta_0 + \frac{\vartheta_1 d}{\vartheta_2 + d}, \sigma_1^2)$ and (μ, σ_2^2) , respectively, then the locally D -optimal design for the dose response model with an active control allocates 22.2% of the patients to each of the dose levels L , $d^* = \frac{R(L+\vartheta_2)+L(R+\vartheta_2)}{(L+\vartheta_2)+(R+\vartheta_2)}$ and R of the new drug and 33.3% to the active control.*
- (b) *In the case of negative binomial distributions with probabilities $\pi(d, \theta) = \vartheta_0 + \frac{\vartheta_1 d}{\vartheta_2 + d}$ and μ , the locally D -optimal design for the dose response model with an active control allocates 25% of the patients to each of the dose levels L , d^* and R of the new drug and 25% to the active control, where d^* is the solution of the equation*

$$\frac{2}{d-L} + \frac{2}{d-R} - \frac{\vartheta_0 + \vartheta_1 - 1}{d(\vartheta_0 + \vartheta_1 - 1) + (\vartheta_0 - 1)\vartheta_2} - \frac{2(\vartheta_0 + \vartheta_1)}{\vartheta_0(\vartheta_2 + d) + \vartheta_1 d} - \frac{1}{\vartheta_2 + d} = 0$$

- (c) *In the case of binomial distributions with probabilities $\pi(d, \theta) = \vartheta_0 + \frac{\vartheta_1 d}{\vartheta_2 + d}$ and μ , the locally D -optimal design is of the same form as described in part (b), where d^* is the solution of the equation*

$$\frac{2}{d-L} + \frac{2}{d-R} - \frac{\vartheta_0 + \vartheta_1 - 1}{d(\vartheta_0 + \vartheta_1 - 1) + (\vartheta_0 - 1)\vartheta_2} - \frac{\vartheta_0 + \vartheta_1}{\vartheta_0(\vartheta_2 + d) + \vartheta_1 d} - \frac{2}{\vartheta_2 + d} = 0.$$

- (d) *If Poisson distributions with parameters $\lambda(d, \theta_1) = \vartheta_0 + \frac{\vartheta_1 d}{\vartheta_2 + d}$ and μ are used in (2.1) and (2.2), respectively, then the locally D -optimal is of the same form as described in part (b), where*

$$d^* = \vartheta_2 \frac{4m(L)m(R) - \vartheta_1(Lm(R) + Rm(L)) - \vartheta_0\sqrt{\kappa}}{-4m(L)m(R) - \vartheta_1\vartheta_2(m(R) + m(L)) + (\vartheta_1 + \vartheta_0)\sqrt{\kappa}}$$

and $\kappa = ((\vartheta_2 + L)m(R) + (\vartheta_2 + R)m(L))^2 + 12(\vartheta_2 + L)(\vartheta_2 + R)m(R)m(L)$, $m(d) = \vartheta_0\vartheta_2 + \vartheta_1 d + \vartheta_0 d$.

Example 3.1 Under the assumption of a normal distribution, Dette et al. (2014) determined a D -optimal design for the EMAX dose response model ignoring the effect caused by

estimating the variance. It follows from Theorem 4 in Dette et al. (2014) that the locally D -optimal design allocates 26.6% of the patients to the dose levels L, d^*, R of the new drug and 20% to the active control, respectively, where d^* is defined in 3.2(a). Theorem 3.2 above shows that the design which accounts for the problem of estimating the variances uses the same dose levels but allocates 13.3% more patients to the active control.

Example 3.2 In this example we discuss D -optimal designs for the two clinical trials considered in Section 1.

- (a) We first consider the gouty arthritis example. The primary endpoint is modeled by a negative binomial distribution with parameters r_1 and $\pi(d, \theta_1) = \vartheta_0 + \frac{\vartheta_1 d}{\vartheta_2 + d}$ for the new drug and parameters r_2 and θ_2 for the comparator. The dose range is $[0, 300]$ mg and we obtained from the clinical team the following preliminary information for the unknown parameters: $\vartheta_0 = 0.26$, $\vartheta_1 = 0.73$, $\vartheta_2 = 10.5$, $\sigma_1 = 0.05$ and $\theta_2 = 0.9206$, $\sigma_2 = 0.05$. In addition, $r_1 = r_2 = 10$ are fixed. The D -optimal design is obtained from Theorem 3.2 and depicted in the upper part of Table 1. It allocates 25% of all patients to the active control and 25% of the patients to the dose levels 0, 8.23, 300mg of the new drug, respectively. The standard design actually used in this study allocates 14.3% of the patients to the dose levels 25, 50, 100, 200, 300mg of the new drug and 28.5% of the patients to the active control. To compare these designs we also show in the last column of Table 1 the D -efficiency

$$\text{eff}_D(\xi, \theta) = \frac{\Phi_0(\xi, \theta)}{\Phi_0(\xi_D^*, \theta)} \in [0, 1], \quad (3.1)$$

where ξ_D^* is the locally D -optimal design. We observe that in this example an optimal design improves the standard design substantially. We also observe that the differences between the D -optimal designs calculated under a different distributional assumption are rather small. For this example, the D -optimal design calculated under the assumption of a normal distribution has efficiency 0.98 in the model based on the negative binomial distribution.

- (b) We now consider the acute migraine example, which measured the percentage of patients reporting pain freedom at two hours post-dose. We assume a binomial distribution for this example. The probabilities of success are $\pi(d, \theta_1) = \vartheta_0 + \frac{\vartheta_1 d}{\vartheta_2 + d}$ for the new compound (where the dose level varies in the interval $[0, 200]$ mg) and θ_2 for the active control. The sample sizes are $n_1 = 517$ and $n_2 = 100$ and the preliminary information obtained from the clinical team is given by $\vartheta_0 = 0.098$, $\vartheta_1 = 0.2052$, $\vartheta_2 = 12.3$, $\sigma_1 = 0.05$ and $\theta_2 = 0.2505$, $\sigma_2 = 0.05$. The locally D -optimal designs under a normal and binomial distribution assumption are listed in the lower part of Table 1. The design actually used for this study allocated 21, 5, 7, 10, 10, 11, 10, 10% of the patients

distribution	D-optimal design				eff _D
normal	(0,0)	(9.81,0)	(300,0)	(C,1)	0.25
	22.2%	22.2%	22.2%	33.3%	
negative binomial	(0,0)	(8.23,0)	(300,0)	(C,1)	0.11
	25%	25%	25%	25%	
normal	(0,0)	(10.95,0)	(200,0)	(C,1)	0.84
	22.2%	22.2%	22.2%	33.3%	
binomial	(0,0)	(9.05,0)	(200,0)	(C,1)	0.86
	25%	25%	25%	25%	

Table 1: *D*-optimal designs in the two clinical trials discussed in Section 1 under different distributional assumptions. Upper part: gouty arthritis example; lower part: acute migraine example. The last column shows the efficiencies of the designs, which were actually used in the study.

to the dose levels 0, 2.5, 5, 10, 20, 50, 100, 200mg of the new drug and 16% of the patients to the active control, respectively. The second column of Table 1 displays its efficiencies relative to the proposed designs and again a substantial improvement can be observed under both distributional assumptions. For this example, the *D*-optimal design calculated under the assumption of a normal distribution has also efficiency 0.98 in the model based on the binomial distribution.

4 Optimal designs for estimating the target dose

In this section we investigate the problem of constructing locally optimal designs for estimating the treatment effect of the active control and the target dose, that is the smallest dose of the new compound which achieves the same treatment effect as the active control. For this purpose we consider a dose range of the form $\mathcal{D} = [L, R]$ and introduce the notation

$$\mathbb{E}_{\theta_1}[Y_{ij}|d_i] = \eta(d_i, \theta_1) \quad (j = 1, \dots, n_{1i}, \quad i = 1, \dots, k) \quad (4.1)$$

$$\mathbb{E}_{\theta_2}[Z_i] = \Delta \quad (i = 1, \dots, n_2) \quad (4.2)$$

for the expected values of responses corresponding to the new drug (for dose level d_i) and the active control, respectively. We assume (for simplicity) that the function η in (4.1) is strictly increasing in $d \in \mathcal{D}$ and that $d^*(\theta) = \eta^{-1}(\Delta, \theta_1)$, is an element of the dose range $\mathcal{D} = [L, R]$ for the new drug. Note that the expectation Δ in (4.2) is a function of the s_2 -dimensional parameter θ_2 , say $\Delta = k(\theta_2)$. Consequently, a natural estimate of d^* is given by $\hat{d}^* = d^*(\hat{\theta}) = \eta^{-1}(\hat{\Delta}, \hat{\theta}_1)$, where $\hat{\Delta} = k(\hat{\theta}_2)$ and $\hat{\theta} = (\hat{\theta}_1^T, \hat{\theta}_2^T)^T$ denotes the vector of the maximum likelihood estimates of the parameter θ_1 and θ_2 in models (2.1) and (2.2), respectively. Standard calculations show that the variance of this estimator is approximately given by

$$\text{Var}(d^*(\hat{\theta})) \approx \frac{1}{N} \psi(\xi, \theta), \quad (4.3)$$

where the function ψ is defined by

$$\psi(\xi, \theta) = \frac{1}{1-\omega_{k+1}} \left(\frac{\partial}{\partial \theta_1} d^*(\theta) \right)^T M_1^-(\tilde{\xi}, \theta_1) \left(\frac{\partial}{\partial \theta_1} d^*(\theta) \right) + \frac{1}{\omega_{k+1}} \left(\frac{\partial}{\partial \theta_2} d^*(\theta) \right)^T I_2^-(\theta_2) \left(\frac{\partial}{\partial \theta_2} d^*(\theta) \right), \quad (4.4)$$

$\tilde{\xi}$ denotes the design for the new drug induced by the design ξ , see (2.18), and $M_1^-(\tilde{\xi}, \theta_1)$ and $I_2^-(\theta_2)$ are generalized inverses of the information matrices $M_1(\tilde{\xi}, \theta_1)$ and $I_2(\theta_2)$, respectively. Following Dette et al. (2014), we call a design ξ_{AC}^* locally AC-optimal design (for Active Control) if $\frac{\partial}{\partial \theta_1} d^*(\theta) \in \text{Range}(M_1(\tilde{\xi}, \theta_1))$, $\frac{\partial}{\partial \theta_2} d^*(\theta) \in \text{Range}(I_2(\theta_2))$ and if ξ_{AC}^* minimizes the function $\psi(\xi, \theta)$ among all designs satisfying this estimability condition. Note that the criterion (4.4) corresponds to a ϕ_{-1} -optimal design for estimating the parameter $K^T \theta$ in a dose response model with an active control, where the matrix K is given by $K = \left(\left(\frac{\partial}{\partial \theta_1} d^*(\theta) \right)^T, \left(\frac{\partial}{\partial \theta_2} d^*(\theta) \right)^T \right)^T$. In particular, Theorem 2.2 is applicable and locally AC-optimal designs can be derived from the corresponding optimal designs for model (2.1). The following result provides an alternative representation of the criterion (4.4) in the case $s_2 = 1$. As a consequence the design $\tilde{\xi}$ required in Theorem 2.2 is a locally \tilde{c} -optimal design in model (2.1) for a specific vector \tilde{c} , i.e. the design minimizing $\tilde{c}^T M_1^-(\tilde{\xi}, \theta_1) \tilde{c}$, where $\tilde{c} = \frac{\partial}{\partial \theta_1} \eta(d^*, \theta_1)$.

Theorem 4.1 *In the case $s_2 = 1$, the function in (4.4) can be represented as*

$$\psi(\xi, \theta) = \frac{\left(\frac{\partial}{\partial \theta_2} d^*(\theta) \right)^2}{\left(\frac{\partial}{\partial \theta_2} k(\theta_2) \right)^2} \left\{ \frac{1}{1-\omega_{k+1}} \left(\frac{\partial}{\partial \theta_1} \eta(d^*, \theta_1) \right)^T M_1^-(\tilde{\xi}, \theta_1) \left(\frac{\partial}{\partial \theta_1} \eta(d^*, \theta_1) \right) + \left(\frac{\partial}{\partial \theta_2} k(\theta_2) \right)^2 \frac{I_2^-(\theta_2)}{\omega_{k+1}} \right\}.$$

In the following discussion we determine locally AC-optimal designs for several nonlinear regression models accounting for an active control by minimizing the criterion (4.4).

4.1 Some explicit results for two-dimensional models

In this section we present some examples illustrating different structures of locally AC optimal designs. For this purpose we consider the situation where the Fisher information matrix $I_1(d, \theta_1)$ defined in (2.5) is of the form

$$I_1(d, \theta_1) = \begin{pmatrix} f(d, \theta_1) f^T(d, \theta_1) & \mathbf{0} \\ \mathbf{0} & \Sigma(\theta_1) \end{pmatrix} \in \mathbb{R}^{s_1 \times s_1} \quad (4.5)$$

where $f(d, \theta_1) = (f_1(d, \theta_1), f_2(d, \theta_1))^T$ denotes a two-dimensional vector and $\Sigma(\theta_1)$ a $(s_1 - 2) \times (s_1 - 2)$ matrix, which does not depend on the dose level. By Theorem 2.2 the locally AC-optimal design can be determined from the design $\tilde{\xi}^*$ which minimizes the expression

$$\tilde{c}^T M_1^-(\tilde{\xi}, \theta_1) \tilde{c} \quad (4.6)$$

in the class of all designs defined on the dose range \mathcal{D} for the new drug, where the vector is given by $\tilde{c} = (\frac{\partial}{\partial \theta_1} d^*(\theta))^T$. Because of the block structure of the Fisher information in (4.5) (with a lower block not depending on the dose level) we may assume without loss of generality that $s_1 = 2$, that is

$$M_1(\tilde{\xi}, \theta_1) = \int_{\mathcal{D}} f(d, \theta_1) f^T(d, \theta_1) d\tilde{\xi}(d). \quad (4.7)$$

By Elfving's theorem [see Elfving (1952)] a design $\tilde{\xi}^*$ with weights \tilde{w}_i^* at the points d_i^* ($i = 1, \dots, k$) minimizes the expression (4.6) if and only if there exists a constant $\gamma > 0$ and $\varepsilon_1, \dots, \varepsilon_k \in \{-1, 1\}$, such that the point $\gamma \tilde{c}$ is a boundary point of the Elfving set

$$\mathcal{R} = \text{conv} \left\{ \varepsilon f(d, \theta_1) \mid d \in \mathcal{D}, \varepsilon \in \{-1, 1\} \right\} \quad (4.8)$$

and the representation $\gamma \tilde{c} = \sum_{i=1}^k \varepsilon_i \tilde{w}_i^* f(d_i^*, \theta_1)$ is valid. Note that $\mathcal{R} = \text{conv}\{\mathcal{C} \cup (-\mathcal{C})\}$, where the curve \mathcal{C} is defined by $\mathcal{C} = \{f(d, \theta_1) \mid d \in \mathcal{D}\}$. The structure of the Elfving set \mathcal{R} depends sensitively on the distributional assumptions and we now consider several examples in the Michaelis-Menten model.

Example 4.1

Assume that the dependence on the dose in model (2.1) is described by the Michaelis-Menten model, then the vector f in (4.7) has the form $v(d, \theta_1) (\frac{d}{\vartheta_2 + d}, -\frac{\vartheta_1 d}{(\vartheta_2 + d)^2})^T$, where the function v varies with the distributional assumption.

- (a) In the case of normal distributed responses we have $v(d, \theta_1) = 1$ and it follows by an obvious generalization of Theorem 4.1, that we have to consider a \tilde{c} -optimal design problem in model (2.1), where the vector \tilde{c} is now given by $\tilde{c} = \frac{\partial}{\partial \vartheta} \eta(d^*, \vartheta) = (\frac{d^*}{\vartheta_2 + d^*}, -\frac{\vartheta_1 d^*}{(\vartheta_2 + d^*)^2})^T$. From the left panel of Figure 2 we observe that the line $\{\gamma \tilde{c} \mid \gamma > 0\}$ intersects the boundary of the Elfving set \mathcal{R} at some point $\mathcal{C} \cup (-\mathcal{C})$, whenever $L \leq x^* \leq d^* < R$, where

$$x^* = L \vee \frac{\sqrt{2}R^2\vartheta_2 + (\sqrt{2}-1)R\vartheta_2^2}{2R^2 + 4R\vartheta_2 + \vartheta_2^2}.$$

A typical situation is shown for the vector \tilde{c}_2 in the left panel of Figure 2 for $\vartheta_1 = \vartheta_2 = 2, \mathcal{D} = [0.1, 50]$. Consequently, Elfving's theorem shows that a one-point design minimizes (4.6) in this case. An application of Theorem 2.2 yields that the locally AC-optimal design which allocates $\frac{\sigma_1}{\sigma_1 + \sigma_2} 100\%$ of the patients to dose level $d^* = \eta^{-1}(\Delta, \vartheta)$ for the new drug and the remaining patients to the active control. On the other hand, if $L < d^* \leq x^* < R$, the line $\{\gamma \tilde{c} \mid \gamma > 0\}$ does not intersect the set $\mathcal{C} \cup (-\mathcal{C})$ at the boundary of the Elfving set \mathcal{R} and the situation is more complicated. A typical situation for this case is shown for the vector \tilde{c}_1 and the locally AC-optimal design allocates $\rho \tilde{w}_1 100\%$, $\rho \tilde{w}_2 100\%$ of the patients to dose levels x^* and R of the new drug,

where $\rho = \frac{\sqrt{\delta}\sigma_1}{\sqrt{\delta}\sigma_1 + \sigma_2}$ and the remaining patients to the active control, where

$$\tilde{\omega}_1 = \frac{v(R, \theta_1)R(R-d^*)(\vartheta_2+x^*)^2}{v(R, \theta_1)R(R-d^*)(\vartheta_2+x^*)^2 + v(x^*, \theta_1)x^*(x^*-d^*)(\vartheta_2+R)^2}, \quad (4.9)$$

$$\tilde{\omega}_2 = 1 - \tilde{\omega}_1, \delta = \tilde{c}^T M_1^{-1}(\tilde{\xi}^*, \theta_1)\tilde{c} \text{ and } d^* = \eta^{-1}(\Delta, \theta_1).$$

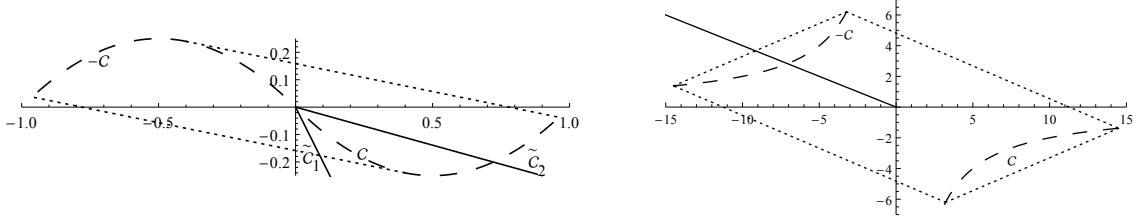


Figure 2: The Elfving set (4.8) in model (2.1), where the expected response is given by the Michaelis-Menten model. Left panel: normal distribution. Right panel: Negative-binomial distribution.

- (b) As a further example consider the Michaelis Menten model for the probability of a negative binomial distributed response. We have $s_1 = 2, s_2 = 1, \pi(d, \theta_1) = \frac{\vartheta_1 d}{\vartheta_2 + d}, \tilde{c} = \frac{\partial}{\partial \theta_1} \eta(d^*, \theta_1) = \frac{r_1}{\vartheta_1 d^*} (-\frac{\vartheta_2 + d^*}{\vartheta_1}, 1)^T$ and the function v is given by $v(d, \theta_1) = \sqrt{\frac{r_1(d+\vartheta_2)^3}{d^2 \vartheta_1^2(d(1-\vartheta_1)+\vartheta_2)}}$. A corresponding Elfving set is depicted in the right panel of Figure 2 for $\vartheta_1 = 1, \vartheta_2 = 0.5, \mathcal{D} = [0, 10]$ and the locally AC-optimal design is always supported at three points. A straightforward calculation shows that the locally AC-optimal design allocates $\rho \tilde{\omega}_1 100\%, \rho \tilde{\omega}_2 100\%$ of the patients to the dose levels L, R for the new drug, where $\rho = \frac{\delta \theta_2^2 - \sqrt{(1-\theta_2)\delta \theta_2^2 r_2}}{\delta \theta_2^2 - (1-\theta_2)r_2}$ and the remaining patients to the active control, where

$$\tilde{\omega}_1 = \frac{v(R, \theta_1)R(R-d^*)(\vartheta_2+L)^2}{v(R, \theta_1)R(R-d^*)(\vartheta_2+L)^2 + v(L, \theta_1)L(d^*-L)(\vartheta_2+R)^2}, \quad (4.10)$$

$$\tilde{\omega}_2 = 1 - \tilde{\omega}_1, \delta = \tilde{c}^T M_1^{-1}(\tilde{\xi}^*, \theta_1)\tilde{c} \text{ and } d^* = \eta^{-1}(\Delta, \theta_1).$$

- (c) Consider now the Michaelis Menten model for binomial distributed responses. We have $s_1 = 2, s_2 = 1, \pi(d, \theta_1) = \frac{\vartheta_1 d}{\vartheta_2 + d}, \tilde{c} = \frac{\partial}{\partial \theta_1} \pi(d^*, \theta_1)$ and $v(d, \theta_1) = \sqrt{\frac{(d+\vartheta_2)^2}{d \vartheta_1 (d(1-\vartheta_1)+\vartheta_2)}}$. The corresponding Elfving set is depicted in the left panel of Figure 3 for $\vartheta_1 = 1, \vartheta_2 = 0.1, \mathcal{D} = [0.02, 2]$ and we have to distinguish three different cases. We observe that the line $\{\gamma \tilde{c} \mid \gamma > 0\}$ intersects the boundary of the Elfving set \mathcal{R} at some point $\mathcal{C} \cup (-\mathcal{C})$ if and only if $L \leq x_1^* \leq d^* \leq x_2^* \leq R$, where

$$x_1^* = L \vee \frac{\vartheta_2(1-\sqrt{1-\pi(R, \theta_1)})}{2\vartheta_1-1+\sqrt{1-\pi(R, \theta_1)}}, \quad x_2^* = R \wedge \frac{\vartheta_2(1+\sqrt{1-\pi(R, \theta_1)})}{2\vartheta_1-1-\sqrt{1-\pi(R, \theta_1)}}.$$

A typical situation is shown for the vector \tilde{c}_1 in the left panel of Figure 3. Consequently, the same arguments as in the previous examples show that in this case the locally

AC-optimal design allocates $\rho 100\%$ of the patients to the dose level d^* of the new drug, where $\rho = \frac{\delta - \sqrt{\delta(1-\theta_2)\theta_2}}{\delta - (1-\theta_2)\theta_2}$ and the remaining patients to the active control, where $\delta = \tilde{c}^T M_1^{-1}(\tilde{\xi}^*, \theta_1) \tilde{c}$ and $d^* = \eta^{-1}(\Delta, \theta_1)$.

On the other hand, if $L < d^* \leq x_1^*$, the locally AC-optimal design allocates $\rho \tilde{\omega}_{11} 100\%$, $\rho(1 - \tilde{\omega}_{11}) 100\%$ of the patients to dose levels x_1^* and R of the new drug and the remaining patients to the active control, where $\tilde{\omega}_{11}$ is of the form (4.9) with $x^* = x_1^*$. A typical situation is shown for the vector \tilde{c}_2 . The case $L \leq x_2^* \leq d^* \leq R$ corresponds to the vector \tilde{c}_3 . Here the locally AC-optimal design allocates $\rho \tilde{\omega}_{21} 100\%$, $\rho(1 - \tilde{\omega}_{21}) 100\%$ of the patients to dose levels x_2^* and R of the new drug and the remaining patients to the active control, where with $L = x_2^*$ $\tilde{\omega}_{21}$ is of the form (4.10) and $d^* = \eta^{-1}(\Delta, \theta_1)$.

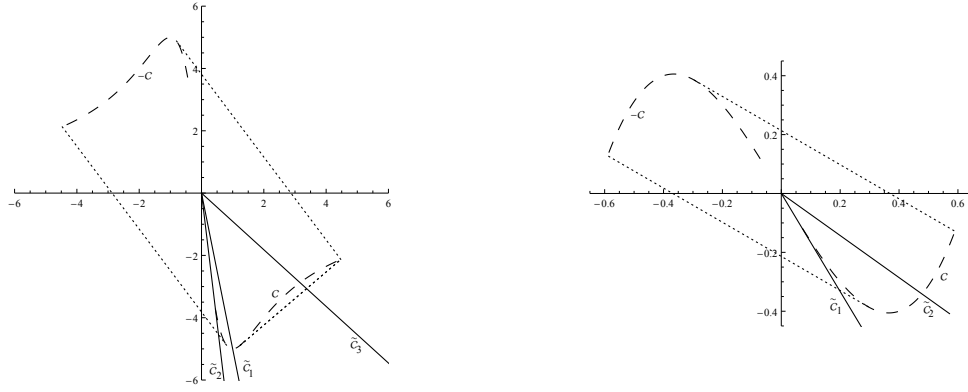


Figure 3: *The Elfving set (4.8) in model (2.1), where the expected response is given by the Michaelis-Menten model. Left panel: binomial distribution. Right panel: Poisson distribution.*

- (d) Finally we consider the case of Poisson distributed responses. We have $s_1 = 2, s_2 = 1, \lambda(d, \theta_1) = \frac{\vartheta_1 d}{\vartheta_2 + d}, v(d, \theta_1) = 1/\sqrt{\frac{\vartheta_1 d}{\vartheta_2 + d}}$ and by Theorem 4.1 we have to solve a \tilde{c} -optimal design problem with $\tilde{c} = \frac{\partial}{\partial \theta_1} \lambda(d^*, \theta_1) = (\frac{d^*}{\vartheta_2 + d^*}, -\frac{\vartheta_1 d^*}{(\vartheta_2 + d^*)^2})^T$. It is easy to see that the line $\{\gamma \tilde{c} | \gamma > 0\}$ intersects the boundary of the Elfving set \mathcal{R} at some point $\mathcal{C} \cup (-\mathcal{C})$ if and only if $L \leq x^* \leq d^* < R$, where $x^* = L \vee \frac{R\vartheta_2}{3R+4\vartheta_2}$ (see the right panel of Figure 3 for $\vartheta_1 = 2.5, \vartheta_2 = 1.5, \mathcal{D} = [0.02, 10]$ and the vector \tilde{c}_2). Consequently, the same arguments as in the previous examples show that in this case the locally AC-optimal design allocates $\rho 100\%$ of the patients to dose levels d^* of the new drug, where $\rho = \frac{\sqrt{\delta}}{\sqrt{\delta} + \sqrt{\theta_2}}$ and the remaining patients to the active control, where $\delta = \frac{d^* \vartheta_1}{\vartheta_2 + d^*}$ and $d^* = \lambda^{-1}(\Delta, \theta_1)$.

On the other hand, if $L < d^* \leq x^* < R$, the locally AC-optimal design allocates $\rho \tilde{\omega}_1 100\%$, $\rho(1 - \tilde{\omega}_1) 100\%$ of the patients to dose levels x^* and R of the new drug and the remaining patients to the active control, where $\tilde{\omega}_1$ is of the form (4.9) with $\delta = (\frac{\partial}{\partial \theta_1} \eta(d^*, \theta_1))^T M_1^{-1}(\tilde{\xi}^*, \theta_1) (\frac{\partial}{\partial \theta_1} \eta(d^*, \theta_1))$. A typical situation is shown for the vector \tilde{c}_1 in the right panel of Figure 3.

4.2 Locally AC-optimal designs in the EMAX model

Explicit expressions for the AC-optimal designs in the EMAX model are very complicated and for the sake of brevity and better illustration we conclude this paper discussing AC-optimal designs for the two data examples from Section 1.

We begin with the gouty arthritis clinical trial where we use the same prior information as in Example 3.2. AC-optimal designs under the assumption of a normal and negative binomial distribution can be found in the upper part of Table 2. For example under the assumption of normal distributed endpoints, the AC-optimal design allocates almost half of the patients to the dose level 101.06mg and the rest to the active control. In order to compare the standard design introduced in Example 3.2 we display in the right column the efficiency

$$\text{eff}_{AC}(\xi, \theta) = \frac{\psi(\xi_{AC}^*, \theta)}{\psi(\xi, \theta)} \in [0, 1], \quad (4.11)$$

where $\psi(\xi, \theta)$ is defined in (2.13) and ξ_{AC}^* is the locally AC-optimal design. For example,

distribution	AC-optimal design			eff_{AC}
normal	(101.06, 0)	(C, 1)		0.66
	49.99%	50.01%		
negative binomial	(5.44, 0)	(300, 0)	(C, 1)	0.48
	7.6%	35.6%	56.8%	
normal	(35.739, 0)	(C, 1)		0.48
	49.99%	50.01%		
binomial	(0, 0)	(200, 0)	(C, 1)	0.47
	7.34%	41.95%	50.71%	

Table 2: AC-optimal designs in the two examples from section 1 under different distributional assumptions. Upper part: gouty arthritis example, with target dose $d^* = 100\text{mg}$; lower part: acute migraine example, with target dose $d^* = 35.6\text{mg}$. The last column shows the efficiencies of the designs, which were actually used in the study.

the efficiency of the standard design for estimating the target dose under the assumption of a normal or negative binomial distribution is 66% and 48%, respectively.

The second trial is the one in treating migraine and again we use the prior information from Example 3.2. AC-optimal designs for normal and binomial distributed responses can be found in the lower part of Table 2. The efficiencies of the standard design are given by 48% and 47% under the assumption of a normal and binomial distribution, respectively.

Acknowledgements The authors would like to thank Martina Stein, who typed parts of this manuscript with considerable technical expertise. This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation (DFG) and by the National Institute

Of General Medical Sciences of the National Institutes of Health under Award Number R01GM107639. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

5 Conclusions

In this paper the optimal design problem for active controlled dose finding studies is considered. Sufficient conditions are provided such that the optimal design for a dose finding study with no active control can also be used for the model with an active control. Our results apply to general optimality criteria and distributional assumptions. In particular they are applicable in models with discrete responses, which appeared recently in two of our consulting projects. In several examples it is demonstrated that the optimal designs may depend sensitively on the distributional assumptions. In the clinical trials under consideration these differences were less visible for D -optimal designs. However, in the problem of estimating the target dose (i.e. the smallest dose of the new compound which achieves the same treatment effect as the active control), the differences are more substantial, and an optimal design calculated under a "wrong" distributional assumption (i.e. a normal distribution) might be inefficient, if it used in a different model (i.e. a Binomial model).

References

- Bornkamp, B., Bretz, F., Dette, H., and Pinheiro, J. (2011). Response-adaptive dose-finding under model uncertainty. *Annals of Applied Statistics*, 5(2B):1611–1631.
- Bornkamp, B., Bretz, F., Dmitrienko, A., Enas, G., Gaydos, B., Hsu, C.-H., König, F., Krams, M., Liu, Q., Neuenschwander, B., Parke, T., Pinheiro, J. C., Roy, A., Sax, R., and Shen, F. (2007). Innovative approaches for designing and analyzing adaptive dose-ranging trials. *Journal of Biopharmaceutical Statistics*, 17:965–995.
- Bretz, F., Hsu, J. C., Pinheiro, J. C., and Liu, Y. (2008). Dose finding - a challenge in statistics. *Biometrical Journal*, 50(4):480–504.
- Brown, L. D. (1986). *Fundamentals of statistical exponential families with applications in statistical decision theory*. Institute of Mathematical Statistics, Hayward, CA.
- Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304.
- Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Annals of Mathematical Statistics*, 24:586–602.
- Dette, H. (1997). Designing experiments with respect to "standardized" optimality criteria. *Journal of the Royal Statistical Society, Ser. B*, 59:97–110.

- Dette, H., Bretz, F., Pepelyshev, A., and Pinheiro, J. C. (2008). Optimal designs for dose finding studies. *Journal of the American Statistical Association*, 103(483):1225–1237.
- Dette, H., Kiss, C., Benda, N., and Bretz, F. (2014). Optimal designs for dose finding studies with an active control. *Journal of the Royal Statistical Society, Ser. B*, 76(1):265–295.
- Dragalin, V., Hsuan, F., and Padmanabhan, S. K. (2007). Adaptive designs for dose-finding studies based on sigmoid e_{\max} model. *Journal of Biopharmaceutical Statistics*, 17(6):1051–1070.
- Elfving, G. (1952). Optimal allocation in linear regression theory. *Annals of Statistics*, 23:255–262.
- EMA (2005). CPMP guideline on the choice of the non-inferiority margin. doc. ref. [emea/cmp/ewp/2158/99](http://www.ema.europa.eu/ema). Technical report, Available at <http://www.ema.europa.eu/ema>.
- EMA (2006). CPMP guideline on clinical investigation of medicinal products for the treatment of multiple sclerosis. doc. ref. [cmp/ewp/561/98](http://www.ema.europa.eu/ema). Technical report, Available at <http://www.ema.europa.eu/ema>.
- EMA (2011). CHMP guideline on medicinal products for the treatment of insomnia. doc. ref. [ema/chmp/16274/2009](http://www.ema.europa.eu/ema). Technical report, Available at <http://www.ema.europa.eu/ema>.
- Fang, X. and Hedayat, A. S. (2008). Locally D-optimal designs based on a class of composed models resulted from blending Emax and one-compartment models. *Annals of Statistics*, 36:428–444.
- Fedorov, V. V. and Leonov, S. L. (2001). Optimal design of dose response experiments: A model-oriented approach. *Drug Information Journal*, 35:1373–1383.
- Ford, I., Torsney, B., and Wu, C. F. J. (1992). The use of canonical form in the construction of locally optimum designs for nonlinear problems. *Journal of the Royal Statistical Society, Ser. B*, 54:569–583.
- Helms, H., Benda, N., and Friede, T. (2014a). Point and interval estimators of the target dose in clinical dose-finding studies with active control. *Journal of Biopharmaceutical Statistics (Early View)*.
- Helms, H., Benda, N., Zinserling, P., Kneib, T., and Friede, T. (2014b). Spline-based procedures for dose-finding studies with active control. *Statistics in medicine (in press)*.
- ICH (1994). Ich Topic E 4. Dose response information to support drug registration. Technical report, Available at <http://www.ema.europa.eu/ema>.
- Imhof, L. A. (2001). Maximin designs for exponential growth models and heteroscedastic polynomial models. *Annals of Statistics*, 29:561–576.
- Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Annals of Statistics*, 2:849–879.
- Krewski, D., Smythe, R., and Fung, K. Y. (2002). Optimal designs for estimating the effective dose in development toxicity experiments. *Risk Analysis*, 22(6):1195–1205.
- Miller, F., Guilbaud, O., and Dette, H. (2007). Optimal designs for estimating the interesting part of a dose-effect curve. *Journal of Biopharmaceutical Statistics*, 17:1097–1115.
- Pinheiro, J., Bretz, F., and Branson, M. (2006). Analysis of dose-response studies: Modeling

- approaches. In Ting, N., editor, *Dose Finding in Drug Development*, pages 146–171. Springer-Verlag, New York.
- Pronzato, L. and Walter, E. (1985). Robust experimental design via stochastic approximation. *Mathematical Biosciences*, 75:103–120.
- Pukelsheim, F. (2006). *Optimal Design of Experiments*. SIAM, Philadelphia.
- Pukelsheim, F. and Rieder, S. (1992). Efficient rounding of approximate designs. *Biometrika*, 79:763–770.
- Rasch, D. (1990). Optimum experimental design in nonlinear regression. *Communications in Statistics: Theory Methods.*, 19(12):4789–4806.
- Ruberg, S. J. (1995). Dose response studies I. Some design considerations. *Journal of Biopharmaceutical Statistics*, 5:1–14.
- Ting, N. (2006). *Dose Finding in Drug Development*. Springer, New York.
- Wu, Y., Fedorov, V. V., and Propert, K. (2005). Optimal design for dose response using beta distributed responses. *Journal of Biopharmaceutical Statistics*, 15(5):753–751.
- Zhu, W. and Wong, W. K. (2000). Multiple-objective designs in a dose-response experiment. *Journal of Biopharmaceutical Statistics*, 10(1):1–14.

6 Appendix: proofs

Proof of Theorem 2.1. Assume that the matrix K is a blockdiagonal matrix of the form (2.16). Observing the representations (2.10), (2.16) and noting that the expression $K^T G K$ is independent of the choice of the generalized inverse of the matrix $M(\xi, \theta)$, we obtain

$$K^T M^-(\xi, \theta) K = \begin{pmatrix} (1 - w_{k+1})^{-1} K_{11}^T M_1^-(\tilde{\xi}, \theta_1) K_{11} & 0 \\ 0 & w_{k+1}^{-1} K_{22}^T I_2^-(\theta_2) K_{22} \end{pmatrix}.$$

In the case $p \neq 0, -\infty$ this gives for the criterion ϕ_p in (2.13) the representation

$$\begin{aligned} \phi_p(\xi) &= \left(\frac{1}{t} \sum_{i=1}^t \lambda_i^{-p} (K^T M^-(\xi, \theta) K) \right)^{\frac{1}{p}} \\ &= \left(\frac{1}{t} \left\{ (1 - w_{k+1})^p \sum_{i=1}^{t_1} \lambda_i^{-p} (K_{11}^T M_1^-(\tilde{\xi}, \theta_1) K_{11}) + w_{k+1}^p \sum_{i=1}^{t_2} \lambda_i^{-p} (K_{22}^T I_2^-(\theta_2) K_{22}) \right\} \right)^{\frac{1}{p}} \\ &= \left(\frac{(1 - w_{k+1})^p t_1}{t} (\tilde{\phi}_p(\tilde{\xi}))^p + \frac{w_{k+1}^p}{t} \text{tr}((K_{22}^T I_2^-(\theta_2) K_{22})^{-p}) \right)^{\frac{1}{p}}, \end{aligned} \tag{5.1}$$

where $\lambda_1(A), \dots, \lambda_n(A)$ denote the eigenvalues of a matrix A , $t = t_1 + t_2$ and the function $\tilde{\phi}_p$ is defined in (2.17).

Now it is easy to see that the function ϕ_p is an increasing function of $\tilde{\phi}_p(\tilde{\xi})$. Consequently, the locally ϕ_p -optimal design problem for the dose response model with an active control can be solved by determining a design $\tilde{\xi}_p^*$ which maximizes the criterion (2.17) in a first step. If $\phi^* = \tilde{\phi}_p(\tilde{\xi}_p^*) = \max_{\tilde{\xi}} \tilde{\phi}_p(\tilde{\xi})$ denotes the optimal value for this criterion, it remains to maximize the function ϕ_p in (5.1) with respect to the weight w_{k+1} assigned to the active control, which gives the expression (2.19) and proves the assertion for the case $p \neq 0, -\infty$. The remaining cases $p = 0$ and $p = -\infty$ are proved similarly and the details are omitted for the sake of brevity. \square

Proof of Theorem 3.1 The proof of part (a) has been given in Section 3. For the remaining cases we restrict ourselves to the case of the Poisson distribution for which the Fisher information in model (2.1) is given by

$$I_1(d, \theta_1) = \frac{d}{\vartheta_1(\vartheta_2 + d)} \begin{pmatrix} 1 & -\frac{\vartheta_1}{\vartheta_2 + d} \\ -\frac{\vartheta_1}{\vartheta_2 + d} & \frac{\vartheta_1^2}{(\vartheta_2 + d)^2} \end{pmatrix} \quad (5.2)$$

[see equation (2.8)]. All other cases are treated similarly. By Corollary 2.1 the D -optimal design can be obtained from the D -optimal design $\tilde{\xi}^*$ in a regression model with Fisher information (5.2). If $M_1(\tilde{\xi}, \theta_1) = \int_{\mathcal{D}} I_1(d, \theta_1) d\tilde{\xi}(d)$ denotes an information matrix of a design $\tilde{\xi}$ in this model, then $\tilde{\xi}^*$ is D -optimal if and only if the inequality $\text{tr}(I_1(d, \theta_1) M_1^{-1}(\tilde{\xi}^*, \theta_1)) \leq 2$ holds for all $d \in \mathcal{D}$ (see Lemma 2.1). Moreover, there must be equality at the support points of the design $\tilde{\xi}^*$. It is easy to see that this inequality is equivalent to an inequality of the form $P_3(d) \leq 0$ where P_3 is a polynomial of degree 3 with $P(0) < 0$. A straightforward argument now shows that $\tilde{\xi}^*$ has exactly two support points $d_1^* > 0$ and $d_2^* = R$. Consequently, the D -optimal design $\tilde{\xi}_1^*$ for the regression model with information matrix (5.2) has equal masses at the points d_1^* and R , where d_1^* maximizes the function

$$f(d) = \frac{R(R-d)^2 d}{4(R + \vartheta_2)^3 (\vartheta_2 + d)^3}$$

in the interval $[L, R]$, that is $d_1^* = L \vee \frac{\vartheta_2 R}{3\vartheta_2 + 2R}$. The assertion now follows by an application of Corollary 2.1, observing that $t_1 = 2, t_2 = 2$ in the case under consideration.

Proof of Theorem 4.1 Note that $\Delta = k(\theta_2)$ and that the dose level $d^*(\theta) = \eta^{-1}(\Delta, \theta_1)$ can be defined as the (unique) solution of the equation $F(d, \theta) = k(\theta_2) - \eta(d, \theta_1) = 0$ with respect to d . Consequently, the implicit function theorem shows that the function $\theta \rightarrow d^*(\theta)$

is differentiable with respect to θ with gradient given by

$$\left(\left(\frac{\partial}{\partial \theta_1} d^*(\theta) \right)^T, \frac{\partial}{\partial \theta_2} d^*(\theta) \right)^T = - \left(\frac{\partial}{\partial d} F(d, \theta) \Big|_{d=d^*(\theta)} \right)^{-1} \left(- \left(\frac{\partial}{\partial \theta_1} \eta(d, \theta_1) \right)^T, \frac{\partial}{\partial \theta_2} k(\theta_2) \right)^T \Big|_{d=d^*(\theta)},$$

which implies (comparing the second components) $-\left(\frac{\partial}{\partial d} F(d, \theta) \Big|_{d=d^*(\theta)} \right)^{-1} = \frac{\partial}{\partial \theta_2} d^*(\theta) / \frac{\partial}{\partial \theta_2} k(\theta_2)$.

Altogether this gives for the first component

$$\left(\frac{\partial}{\partial \theta_1} d^*(\theta) \right)^T = \left(\frac{\partial}{\partial d} F(d, \theta) \Big|_{d=d^*(\theta)} \right)^{-1} \left(\frac{\partial}{\partial \theta_1} \eta(d^*, \theta_1) \right)^T = - \frac{\frac{\partial}{\partial \theta_2} d^*(\theta)}{\frac{\partial}{\partial \theta_2} k(\theta_2)} \left(\frac{\partial}{\partial \theta_1} \eta(d^*, \theta_1) \right)^T$$

and the result follows from the representation (2.13).